A Panel Data Analysis: Research & Development Spillover

Werner Müller, Michaela Nettekoven

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 55
May 1998

http://statmath.wu-wien.ac.at/

A Panel Data Analysis: Research & Development Spillover

Werner G. Müller* and Michaela Nettekoven
Dept. of Statistics, Univ. of Economics Vienna, Austria

May 12, 1998

Abstract

Panel data analysis has become an important tool in applied econometrics and the respective statistical techniques are well described in several recent textbooks. However, for an analyst using these methods there remains the task of choosing a reasonable model for the behavior of the panel data. Of special importance is the choice between so-called fixed and random coefficient models. This choice can have a crucial effect on the interpretation of the analyzed phenomenon, which is demonstrated by an application on research and development spillover.

JEL-code: C23.

Keywords: fixed coefficient models, random coefficient models, iterative estimation.

1 Introduction

Recent economic theories consider research and development expenditures as decisive factors for technological progress and economic growth. Coe & Helpman (1995) postulate that factor productivity depends not only upon domestic but as well on foreign R&D capital stocks. They use a panel data approach based upon a fixed regression coefficients specification to test their hypothesis and come to the conclusion that the R&D investments of trade partners may have a strong positive effect on domestic factor productivity. In this paper we show that an alternative panel data model based upon random coefficients, that is perfectly compatible with the data, may lead to different conclusions.

*Dr. Werner G. Müller, Dept. of Statistics, Univ. of Economics and Business Administration, Augasse 2-6, A-1090 Vienna, Austria, Tel. (1) 31336 4258, FAX (1) 31336 711, email: werner.mueller@wu-wien.ac.at

The initial model employed by Coe & Helpman (1995) is as follows:

$$\ln F_i = \alpha_i^0 + \alpha_i^d \ln S_i^d + \alpha_i^f m_i \ln S_i^f, \tag{1}$$

where F denotes total factor productivity, S_i^d domestic R&D capital stock, S_i^f foreign R&D capital stock, and m_i the fraction of imports in GDP. The index i identifies the respective country. By this model it is hypothesized that countries benefit not only from their own capital stock, but also of that of their trading partners (weighted by the amount of trade).

Eventually Coe & Helpman (1995) use the following modification of the above model:

$$\ln F_i = \alpha_i^0 + \alpha^d \ln S_i^d + \alpha^{G7} G7 \ \ln S_i^d + \alpha^f m_i \ln S_i^f. \tag{2}$$

Here, a dummy variable $G7$ is employed to distinguish deviating behavior of the seven largest economies. Furthermore, now the coefficients α^d, α^f, and α^{G7} are kept constant over all countries and so country specific variations are only modeled by the import factors m_i and the intercepts α_i^0. In the following we will again drop the $G7$ dummy for simplification. Although the dummy is significant it has no effect on our comparison and the essence of the conclusions we reach.

The authors use corresponding data from 21 OECD countries and Israel between 1971 and 1990. They confirm the nonstationarity of the time series and consequently interpret (2) as a cointegrating relationship. The parameters are then estimated by OLS from the pooled data set, which produces results like those given in Table 1. From these results it can be seen that the estimated proportionality factor $\hat{\alpha}^f$ for the elasticity of the total factor product with respect to foreign capital stock is significantly positive (even more so in the $G7$ dummy model), and so Coe & Helpman (1995) conclude that a trade partners R&D capital stock has a positive impact for the domestic economy.

parameter	estimate	st.deviation	t-value
α^d	0.109	0.008	13.252
α^f	0.233	0.045	5.154

Table 1: Estimation results by a fixed coefficient model.

2 Random coefficient models

The analysis undertaken in the previous section assumes fixed regression coefficients. Another popular approach in panel data analysis, however, is to regard the coefficients as random drawings from a distribution with a common mean. The task is then to estimate this mean parameter. Based upon (1) we can easily formulate such a random coefficient model for the current context:

$$\ln F_{it} = X_{it}'(\beta + \gamma_i) + u_{it}, \qquad t = 1, \ldots, T \tag{3}$$

where $X'_{it} = (1, \ln S^d_{it}, m_i \ln S^f_{it})$ and $(\beta + \gamma_i)' = (\alpha^0_i, \alpha^d_i, \alpha^f_i) = \beta'_i$. We require the following assumptions:

$$E(\gamma_i) = 0, E(\gamma_i \gamma'_j) = \begin{cases} \Delta & \text{for } i = j \\ 0 & \text{else} \end{cases}, E(\gamma_i X'_{it}) = 0, E(u_i u_j) = \begin{cases} \sigma^2_i I_T & \text{for } i = j \\ 0 & \text{else} \end{cases}.$$

The noise terms thus consist of

$$X'_{it}\gamma_i + u_{it}$$

and their variance covariance matrices Ω_i depend upon Δ and σ^2_i.

To check whether the random coefficient model is in accordance with the data, Hsiao (1986) proposes to test the hypothesis $H_0 : \quad \beta_1 = \beta_2 = \ldots = \beta_N = \beta$ by calculating the test statistic

$$\mathcal{F} = \sum_{i=1}^{N} \frac{(\hat\beta_i - \tilde\beta)' X'_i X_i (\hat\beta_i - \tilde\beta)}{\hat\sigma^2_i},$$

which is asymptotically F-distributed with $K(N-1)$ degrees of freedom; in our case the number of parameters is $K = 3$ and the number of countries is $N = 22$.

Here $\hat\beta_i$ denotes the individual OLS estimates for the i-th country, $\hat\sigma^2_i$ a consistent estimate of the respective error variance, e.g.

$$\hat\sigma^2_i = \frac{\hat u'_i \hat u_i}{T - K},$$

and

$$\tilde\beta = \left(\sum_{i=1}^{N} \frac{X'_i X_i}{\hat\sigma^2_i}\right)^{-1} \left(\sum_{i=1}^{N} \frac{X'_i y_i}{\hat\sigma^2_i}\right).$$

For the given data set this yields $\mathcal{F} = 3378.26$; the critical value for a significance level of 99% is 92.0, and thus the application of a random coefficient model seems sensible.

3 Estimators

Under the given assumptions the best linear unbiased estimator is the so called generalized least squares estimator given by

$$\hat\beta_{GLS} = \left(\sum_{i=1}^{N} X'_i \Omega_i^{-1} X_i\right)^{-1} \left(\sum_{i=1}^{N} X'_i \Omega_i^{-1} y_i\right)$$

where

$$\Omega_i = X_i \Delta X'_i + \sigma^2_i I_T$$

denotes the i-th diagonal block of the error variance covariance matrix.

However, since the Ω_i are unknown in practice, a different, feasible estimator needs to be found. One suggestion goes back to Swamy (1970), who proposes to replace the σ_i^2 by the consistent estimators $\hat{\sigma}_i^2$ and Δ by

$$\hat{\Delta} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\beta}_i - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i\right)\left(\hat{\beta}_i - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i\right)'$$

to subsequently construct the weighted estimator

$$\hat{\beta}_S = \sum_{i=1}^{N} W_i \hat{\beta}_i \tag{4}$$

with weights

$$W_i = \left\{\sum_{i=1}^{N}[\hat{\Delta} + \hat{\sigma}_i^2(X_i'X_i)^{-1}]^{-1}\right\}^{-1}[\hat{\Delta} + \hat{\sigma}_i^2(X_i'X_i)^{-1}]^{-1}.$$

This estimator is consistent and asymptotically efficient. The consistent and efficient estimator in this framework is provided by an iterative procedure first indicated by Fisk (1967):

$$\hat{\beta}_F = \lim_{j\to\infty}\hat{\beta}(j), \tag{5}$$

with

$$\hat{\beta}_i(j+1) = \arg\min_{\beta}\left\{\sum_{t=1}^{T}[\ln F_{it} - X_{it}'\beta]/\hat{\sigma}^2(j) + [\beta - \hat{\beta}(j)]'\hat{\Delta}^{-1}(j)[\beta - \hat{\beta}(j)]\right\},$$

$$\hat{\sigma}^2(j) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[\ln F_{it} - X_{it}'\hat{\beta}_i(j)]^2,$$

$$\hat{\Delta}(j) = \frac{1}{N}\sum_{i=1}^{N}[\hat{\beta}_i(j) - \hat{\beta}(j)][\hat{\beta}_i(j) - \hat{\beta}(j)]',$$

and

$$\hat{\beta}(j) = \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_i(j).$$

This procedure needs to be started at reasonable first guesses $\hat{\beta}_i$ (say from individual OLS regressions) and can be stopped when the numerical optimization routine fails to identify a clear minimum. For an overview of those and other estimators in the random coefficient model see the review paper by Fedorov et al. (1993).

4 Estimation results and conclusions

Both estimators from the previous section were employed to estimate the mean parameters α^0, α^d, and α^f. The results are displayed in Table 2.

parameter	Swamy's estimator $\hat{\beta}_S$ (4) estimate	st.dev.	t-value	Fisk's iterative estimator $\hat{\beta}_F$ (5) estimate	st.dev.	t-value
α^0	-0.005	0.004	1.300	-0.005	0.003	1.424
α^d	0.283	0.067	4.195	0.263	0.073	3.599
α^f	-0.199	0.101	-1.965	-0.182	0.061	3.005

Table 2: Random coefficient model results for the spillover data.

It is remarkable that for both estimators the estimated mean elasticity α^f of the total factor product with respect to foreign capital stock is negative and for estimator (5) even significantly so. This result does not seem very surprising, since most of the individual OLS estimates for this coefficient were negative as well. Here, we thus observe the interesting effect of a significant sign change for a parameter of interest depending upon whether a fixed or random coefficient model specification is employed.

Concluding we believe that the data interpretation by Coe & Helpman (1995) must at least be doubted in the light of results from a different and reasonable model. Coe et al. (1997) employ the same model and analysis technique in investigating R&D spillover between developed and developing countries. Although this data set has not been reanalysed by using a random coefficient model, similar doubts on the validity of the results might be in place.

Acknowledgement

We are most grateful to Thomas Url for suggesting the problem and helpful comments on how to improve the paper.

References

Coe, D.T. & Helpman, E. (1995). International R&D spillovers. *European Economic Review* **39**, 859–887.

Coe, D.T., Helpman, E. & Hoffmaister, A.W. (1997). North-South R&D Spillovers. *The Economic Journal* **107 (January)**, 131–149.

Fedorov, V.V., Hackl, P. & Müller, W.G. (1993). Estimation and experimental design for second kind regression models. *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* **24**, 134–151.

Fisk, P.R. (1967). Models of the second kind in regression analysis. *Journal of the Royal Statistical Society, Series B* **29**, 266–281.

Hsiao, C. (1986). *Analysis of Panel Data.* Cambridge University Press.

Swamy, P.A.V.B. (1970). Efficient inference in a random coefficient regression model. *Econometrica* **38**, 311–323.